

<b>REPORT DOCUMENTATION PAGE</b>				<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
<b>1. REPORT DATE (DD-MM-YYYY)</b> 8 Sep 2016		<b>2. REPORT TYPE</b> Journal Article		<b>3. DATES COVERED (From – To)</b> Jun 2013 – Jun 2015	
<b>4. TITLE AND SUBTITLE</b>  Comparison of Decision-Assist and Clinical Judgment of Experts for Prediction of Lifesaving Interventions				<b>5a. CONTRACT NUMBER</b>	
				<b>5b. GRANT NUMBER</b>	
				<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b> Colin F Mackenzie, Cheng Gao, Peter F Hu, Amechi Anazodo, Hegang Chen, Theresa Dinardo, P. Cristina Imle, Lauren Hartsy, Christopher Stephens, Jay Menaker, Yvette Fouche, Karen Murdock, Samuel Galvagno, Richard Alcorta, Stacy Shackelford, and the ONPOINT Study Group				<b>5d. PROJECT NUMBER</b>	
				<b>5e. TASK NUMBER</b>	
				<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> USAF School of Aerospace Medicine International Expeditionary Education & Training Dept/ETS 2510 Fifth St., Bldg. 840 Wright-Patterson AFB, OH 45433-7913				<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  AFRL-SA-WP-JA-2014-0017	
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b>				<b>10. SPONSORING/MONITOR'S ACRONYM(S)</b>	
				<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b>  Distribution A: Approved for public release; distribution is unlimited					
<b>13. SUPPLEMENTARY NOTES</b> Cleared, 88PA, Case # 88ABW-2014-1775, 17 Apr 2014; Shock. 2015; 43(3):238-243					
<b>14. ABSTRACT</b> Early recognition of hemorrhage during the initial resuscitation of injured patients is associated with improved survival in both civilian and military casualties. We tested a transfusion and lifesaving intervention (LSI) prediction algorithm in comparison with clinical judgment of expert trauma care providers. We collected 15 min of pulse oximeter photoplethysmograph waveforms and extracted features to predict LSIs. We compared this with clinical judgment of LSIs by individual categories of prehospital providers, nurses, and physicians and a combined judgment of all three providers using the Area Under Receiver Operating Curve (AUROC). We obtained clinical judgment of need for LSI from 405 expert clinicians in 135 trauma patients. The pulse oximeter algorithm predicted transfusion within 6 h (AUROC, 0.92; P = 0.003) more accurately than either physicians or prehospital providers and as accurately as nurses (AUROC, 0.76; P = 0.07). For prediction of surgical procedures, the algorithm was as accurate as the three categories of clinicians. For prediction of fluid bolus, the diagnostic algorithm (AUROC, 0.9) was significantly more accurate than prehospital providers (AUROC, 0.62; P = 0.02) and nurses (AUROC, 0.57; P = 0.04) and as accurate as physicians (AUROC, 0.71; P = 0.06). Prediction of intubation by the algorithm (AUROC, 0.92) was as accurate as each of the three categories of clinicians. The algorithm was more accurate (P = 0.03) for blood and fluid prediction than the combined clinical judgment of all three providers but no different from the clinicians in the prediction of surgery (P = 0.7) or intubation (P = 0.8). Automated analysis of 15 min of pulse oximeter waveforms predicts the need for LSIs during initial trauma resuscitation as accurately as judgment of expert trauma clinicians. For prediction of emergency transfusion and fluid bolus, pulse oximetry features were more accurate than these experts. Such automated decision support could assist resuscitation decisions, trauma team, and operating room and blood bank preparations.					
<b>15. SUBJECT TERMS</b> Automated decision-assist, clinical judgment, pulse oximetry, blood transfusion, trauma resuscitation					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  SAR	<b>18. NUMBER OF PAGES</b>  7	<b>19a. NAME OF RESPONSIBLE PERSON</b> Dr. Colin Mackenzie
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER (include area code)</b>

## OPEN

## COMPARISON OF DECISION-ASSIST AND CLINICAL JUDGMENT OF EXPERTS FOR PREDICTION OF LIFESAVING INTERVENTIONS

Colin F Mackenzie,<sup>\*†‡</sup> Cheng Gao,<sup>§</sup> Peter F Hu,<sup>\*†</sup> Amechi Anazodo,<sup>†</sup> Hegang Chen,<sup>||</sup> Theresa Dinardo,<sup>||\*\*</sup> P. Cristina Imle,<sup>†</sup> Lauren Hartsky,<sup>††</sup> Christopher Stephens,<sup>\*†</sup> Jay Menaker,<sup>†\*\*</sup> Yvette Fouche,<sup>\*†</sup> Karen Murdock,<sup>†</sup> Samuel Galvagno,<sup>\*†</sup> Richard Alcorta,<sup>††</sup> Stacy Shackelford,<sup>\*\*††‡‡</sup> and the ONPOINT Study Group

*\*Department of Anesthesiology, †Shock Trauma Anesthesiology Research Center and Charles McMathias National Study Center for Trauma and EMS, and ‡Department of Physiology, University of Maryland School of Medicine; §Departments of Electrical Engineering, Computer Science, University of Maryland, Baltimore County; ||Department of Epidemiology, ||Shock Trauma Nursing, and \*\*Shock Trauma Center, University of Maryland School of Medicine; ††USAF C-STARS Baltimore, Maryland Institute for Emergency Medical Services Systems; and ‡‡Department of Surgery, University of Maryland School of Medicine, Baltimore, Maryland*

Received 7 Aug 2014; first reviewed completed 25 Aug 2014; accepted in final form 17 Oct 2014

**ABSTRACT**—Early recognition of hemorrhage during the initial resuscitation of injured patients is associated with improved survival in both civilian and military casualties. We tested a transfusion and lifesaving intervention (LSI) prediction algorithm in comparison with clinical judgment of expert trauma care providers. We collected 15 min of pulse oximeter photoplethysmograph waveforms and extracted features to predict LSIs. We compared this with clinical judgment of LSIs by individual categories of prehospital providers, nurses, and physicians and a combined judgment of all three providers using the Area Under Receiver Operating Curve (AUROC). We obtained clinical judgment of need for LSI from 405 expert clinicians in 135 trauma patients. The pulse oximeter algorithm predicted transfusion within 6 h (AUROC, 0.92;  $P < 0.003$ ) more accurately than either physicians or prehospital providers and as accurately as nurses (AUROC, 0.76;  $P = 0.07$ ). For prediction of surgical procedures, the algorithm was as accurate as the three categories of clinicians. For prediction of fluid bolus, the diagnostic algorithm (AUROC, 0.9) was significantly more accurate than prehospital providers (AUROC, 0.62;  $P = 0.02$ ) and nurses (AUROC, 0.57;  $P = 0.04$ ) and as accurate as physicians (AUROC, 0.71;  $P = 0.06$ ). Prediction of intubation by the algorithm (AUROC, 0.92) was as accurate as each of the three categories of clinicians. The algorithm was more accurate ( $P < 0.03$ ) for blood and fluid prediction than the combined clinical judgment of all three providers but no different from the clinicians in the prediction of surgery ( $P = 0.7$ ) or intubation ( $P = 0.8$ ). Automated analysis of 15 min of pulse oximeter waveforms predicts the need for LSIs during initial trauma resuscitation as accurately as judgment of expert trauma clinicians. For prediction of emergency transfusion and fluid bolus, pulse oximetry features were more accurate than these experts. Such automated decision support could assist resuscitation decisions, trauma team, and operating room and blood bank preparations.

**KEYWORDS**—Automated decision-assist, clinical judgment, pulse oximetry, blood transfusion, trauma resuscitation

## INTRODUCTION

Early recognition of hemorrhage during initial resuscitation of the injured patient is associated with improved survival in both civilian and military casualties (1, 2). Rapid and effective

performance of lifesaving interventions (LSIs) is needed to increase survival. However, the diagnosis of occult hemorrhage remains challenging, and patients with compensated shock may deceive clinicians until the moment of rapid decompensation is reached and irreversible shock ensues. Errors in decision making rather than in technical skill have been associated with two thirds of hemorrhagic deaths after injury (3), and uncontrolled hemorrhage remains the most common cause of potentially preventable death after both civilian and military trauma (4, 5).

Early identification of the bleeding patient remains an unrealized goal despite many efforts at developing predictive models using clinical information available early after admission (6, 7). Using an alternative approach, we have developed and tested a novel method of identifying the future need for blood transfusion using a real-time decision-support algorithm that is based on automated analysis of select features of the pulse oximeter photoplethysmograph (PPG) waveform. In our early experience, the PPG-derived algorithm predicted the need for blood transfusion with Area Under the Receiver Operating Curve (AUROC) of 0.83 and of massive transfusion with AUROC of 0.94 (8). We hypothesized that the PPG-derived algorithm would predict the need for blood transfusion and other LSIs as accurately as the clinical judgment of expert

Address reprint requests to Colin F. Mackenzie, MB, ChB, FRCA, FCCM, 419 W. Redwood St, Suite 225, Baltimore, MD. E-mail cmack003@umaryland.edu.

This research was funded in part by the US Air Force (USAF) FA8650-11-2-6D01 and USAF FA8650-11-2-6142 and Office Naval Research (ONR) N00014-12-C-0120. The opinions expressed in this article are those of the authors and do not reflect the opinions of USAF or ONR.

ONPOINT Study Group participation in this study: Amechi Anazodo, MB, BS; John Blenko, MD; Hegang Chen, PhD; Theresa Dinardo, MS; Joseph DuBose, MD, FACS; Raymond Fang, MD, FACS; Yvette Fouche, MD; Linda Goetz, MHA, CRNA; Tom Grissom, MD; Victor Giustina, BS; Lauren Hartsky, BA, EMT-B; George Hagegeorge, BS; Anthony Herrera, MS; Peter Hu, PhD; Cris Imle, MS; Matthew Lissauer, MD, FACS; Colin Mackenzie, MBChB, FRCA, FCCM; Jay Menaker, MD; Karen Murdock, DScPT; Sara Murthi, MD, FACS; Mayur Narayan, MD, PhD; Sarah Saccicchio, BS; Thomas Scalea, MD, FACS; Stacy Shackelford, MD, FACS; Robert Sikorski, MD; Lynn Smith, MS; Lynn Stansbury, MD, MPH; Deborah Stein, MD, MPH, FACS; and Chris Stephens, MD.

DOI: 10.1097/SHK.0000000000000288

Copyright © 2015 by the Shock Society

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License, where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

trauma care providers. Both were tested on the same patient cohort during resuscitation of physiologically unstable trauma patients.

## MATERIALS AND METHODS

### Institutional review boards

The study was approved by expedited review of institutional review boards (IRBs) from both the University of Maryland and the US Air Force, and informed consent was obtained for all University of Maryland R Adams Cowley Shock Trauma Center (STC) attending physicians and fellows (MD) and all STC trauma resuscitation unit nurses (RNs). The IRB approved a waiver of the need to document informed consent from individual prehospital providers (PHPs), with approval through each Regional Emergency Medical Services Executive Committee. No unique identifying information was collected on individuals providing clinical judgment. The RNs were asked to identify the level of experience equal to or less than 3 years; MDs, their fellow or attending status and specialty (anesthesiology or surgery); and PHPs, their level of training as an emergency medical technician basic or as a paramedic.

### Survey timing

More than 5,000 direct trauma patient admissions occur annually to the STC where Advanced Trauma Life Support management protocols for trauma patient reception and resuscitation are practiced. The PHP, RN, and MD providers perform as a team of 5 to 10 members to minimize inconsistencies in initial trauma patient management. Clinical judgment to predict the need for LSIs was usually completed within 10 min of the patient's having been admitted to the trauma center. The RN clinical judgment survey was completed by an RN not providing clinical care to the study patient. Clinicians were requested to check yes/no for each of 12 LSIs if they believed this LSI would/would not occur within the next 6 h. No opportunity was provided by serial examinations to allow for reevaluation of these one-time decisions. Face validity, construct validity, and content validity of the survey instrument were assessed *a priori* by administering the survey instrument to 18 expert (i.e., >10 years' experience working at a Level 1 trauma center) physicians and nurses. Internal consistency of the instrument was high, as evidenced by calculation of Cronbach  $\alpha = 0.91$  (9).

### Data collection and inclusion/exclusion criteria

Under a separate IRB approval, continuous vital signs, including PPG waveforms and clinical outcomes, were collected prospectively on sequential trauma patient admissions, 18 years and older, who survived at least 15 min after direct admission for the scene of injury, who had an abnormal shock index more than 0.61 (10) (SI = heart rate [HR]/systolic blood pressure [SBP]) or were categorized as Emergency Medical Services Priority 1, that is, a critically ill or an injured person requiring immediate attention or as unstable patients with life-threatening injury or illness, without available prehospital vital signs.

A signal quality index was applied to filter pulse oximeter signal artifacts. This excluded up to 30% of the available pulse oximeter waveforms as previously described (8). The algorithm predictions showed no differences until more than 50% of signals were rejected. No patients were excluded because of an inability to obtain good quality waveform signals. Fifteen minutes of data were collected after admission to predict LSI occurring 15 min to 6 h later, and the automated analysis included detection of 30 features of the pulse oximeter signal. The algorithms were designed to predict blood transfusion and emergency surgery within 6 h, fluid bolus, and tracheal intubation within the first hour after patient admission. Twelve PPG features quantified the amplitude from peak to valley of the PPG waveform. Nine features were related to total millivolts of the amplitude of the PPG signal, and the remaining three features included the 25th percentile and 75th percentile of the PPG amplitude and the PPG amplitude interquartile range (IQR = 25th – 75th percentile). Nine features were extracted from the percutaneous oxygen saturation (SpO<sub>2</sub>) signal, and nine features were extracted from the HR signal using independently developed software as described elsewhere (8). Admission laboratory data alone and in combination with PPG data were analyzed separately and were the subject of a recent presentation by our group (11).

### Lifesaving interventions

Actual LSIs were recorded by research assistants colocated with the resuscitation team during the first hour of trauma resuscitation unit care. Six-hour outcomes were collected by patient chart review. The clinical judgment predictions and patient outcomes were entered into Access and Excel databases (Microsoft Inc, Redmond, Wash) for analysis. The AUROC were calculated, and De Long's method was used to compare AUROC of clinical judgment and pulse oximeter signal-derived predictions (12). The prediction from each category of clinician (MD, RN, PHP) was compared with the actual patient

record for each of the 12 specified interventions. Because the surveys were anonymous (as approved by IRB for blanket consent for all in-hospital clinicians and PHPs), an individual clinician's response could not be linked to a particular survey. When not all the clinicians agreed, the best clinician judgment for the occurrence of LSI was also combined into a majority vote, where any two providers' correct predictions of LSI were considered 100% correct and any two providers' incorrect predictions of LSI were considered 100% incorrect, so the majority ruled.

### Statistics and algorithm validation

The robustness of the pulse oximeter algorithm versus clinical judgment was assessed using leave-one-out methodology for cross-validation of the model (13). SAS Version 9.2 (SAS Institute, Cary, NC) was used for all statistical calculations, and a value of  $P < 0.05$  was considered statistically significant.

## RESULTS

### Population demographics and incidence of LSIs

Four-hundred five clinical judgment surveys of need for LSIs were completed. Demographics of the 135 trauma patient admissions with complete data available for PPG waveforms and clinical judgment of RN, PHP, and MD are shown in Table 1. In the study cohort, 10 received blood, 10 had surgery, eight had fluid bolus, six were intubated, and two patients died (1.5%). Twenty-two patients had 34 LSIs (14, one LSI; 4, two LSIs; and 4, three LSIs).

### AUROC and validation results

The AUROC for the clinicians' predictions compared with the pulse oximetry algorithm predictions of actual outcomes are shown in Figure 1. The PPG algorithm was specific to each LSI. For blood transfusion within 6 h, the AUROC of PHP prediction (0.60) was significantly lower ( $P < 0.02$ ) than the clinical judgment of RNs (0.76) or MDs (0.70). The algorithm predicted blood transfusion with AUROC 0.92. This was significantly more accurate ( $P < 0.003$ ) than either the MDs or PHPs, however, was not different than the RNs ( $P = 0.07$ ) for prediction of blood transfusion. For prediction of surgical

TABLE 1. Demographic characteristics of 135 enrolled patients

Mean age (SD), years		39.3 (17.36)	
Mean admission Glasgow Coma Scale score (SD)		13.9 (2.56)	
		n	%
Sex			
Male		95	70.4
Female		40	29.6
Injury type			
Blunt		107	79.3
Penetrating		22	16.3
Other (hanging, drowning, bites, crush, etc)		6	4.4
Mechanism of injury			
Motor vehicle-associated		66	48.9
Falls		22	16.3
Interpersonal violence		33	24.4
Other		14	10.4
Disposition at discharge			
Home or institutional care		133	98.5
Died in hospital		3	2.2

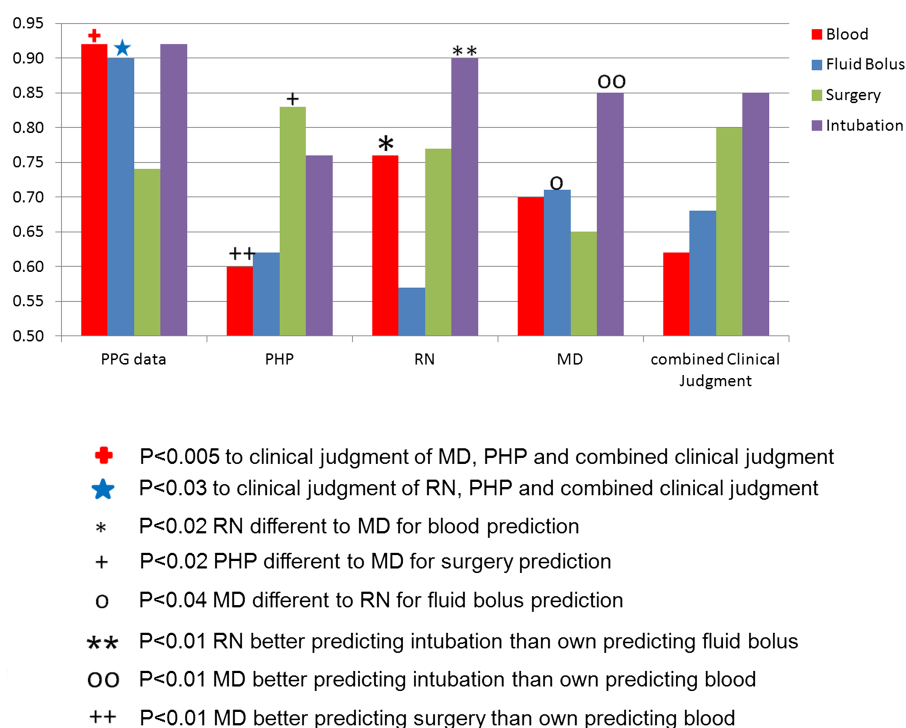


FIG. 1. Comparison of AUROC between photoplethysmograph feature-derived predictions and those of PHPs, RNs, and MDs on the x axis. A combined majority vote clinical judgment (see text for details) is also shown.

interventions within 6 h, the AUROC for the pulse oximeter algorithm (0.74) was not significantly different from the PHP prediction for surgical interventions (AUROC, 0.83). The PHPs were however more accurate ( $P < 0.03$ ) than the judgment of MDs (AUROC, 0.65) but no different from RNs (AUROC, 0.77). For fluid bolus in the first hour, the algorithm predictions provided AUROC of 0.9, significantly more accurate than the PHPs (AUROC, 0.62;  $P < 0.02$ ) and RNs (AUROC, 0.57;  $P < 0.04$ ) but no different from the predictions of MDs (AUROC, 0.71;  $P = 0.06$ ). The algorithm predicted endotracheal intubation with AUROC 0.92, no differently from the clinical judgment of any of the three categories of clinicians. When the predictions of all three categories of clinicians are aggregated as a majority vote, the algorithm predictions were significantly more accurate ( $P < 0.005$ ) for blood transfusion and fluid bolus ( $P < 0.003$ ) but no different for surgery ( $P = 0.7$ ) or intubation ( $P = 0.8$ ). The above results are summarized in Figure 1.

As an illustration of the algorithm predictions versus clinical judgment, Figure 2 shows pulse oximeter PPG signal analysis in a patient who was stabbed multiple times in the abdomen. The PPG algorithm indicated no blood was needed. Clinical judgment of those managing the patient resulted in transfusion of one unit of packed red blood cells before an exploratory laparotomy that showed no evidence of intraperitoneal or significant retroperitoneal injury. The patient had an admission hemoglobin value of 15.7 g/dL, received only this single-unit transfusion, and was discharged home 2 days after admission.

### Sensitivity and specificity and false-negative analyses

Sensitivity (SN = the prediction of true-positive [TP]) and specificity (SP = prediction of true-negative [TN]) were

calculated. Specificity is 1.0 for blood transfusion in this cohort (meaning that the algorithm correctly identified all patients who were predicted to not need transfusion). Sensitivity was 0.7, indicating that the PPG algorithm correctly identified seven of 10 patients who were transfused. As is well known, how SN and SP are reported is a reflection of the “cost” of making a TP or TN decision. As the clinical vignette shown in Figure 2 illustrates, a false-negative rate (FNR = False-negative/No. positive cases) assumes that all the blood that was transfused was indeed indicated. We think that the patient illustrated in Figure 2 did not need blood. If this was an erroneous decision to give blood, then it was (as the algorithm predicted) a TN, not a TP. So, for this reason, we have not reported the FNR.

### Experience impact?

We did not observe an effect of experience level with regard to clinical judgment; there were no significant differences in clinical judgment between fellow and attending physicians, between RNs with experience equal to or less than 3 years, or between emergency medical technician basic and paramedic. In addition, no single category of clinicians consistently provided the most accurate predictions for all LSIs, with MDs, RNs, and PHPs each achieving the most accurate prediction for certain interventions.

### Validation

The results of leave-one-out methodology for validation of clinical judgment of the three categories of clinicians are shown in Table 2. In the prediction of blood transfusion, the algorithm still outperforms the three expert providers, with a

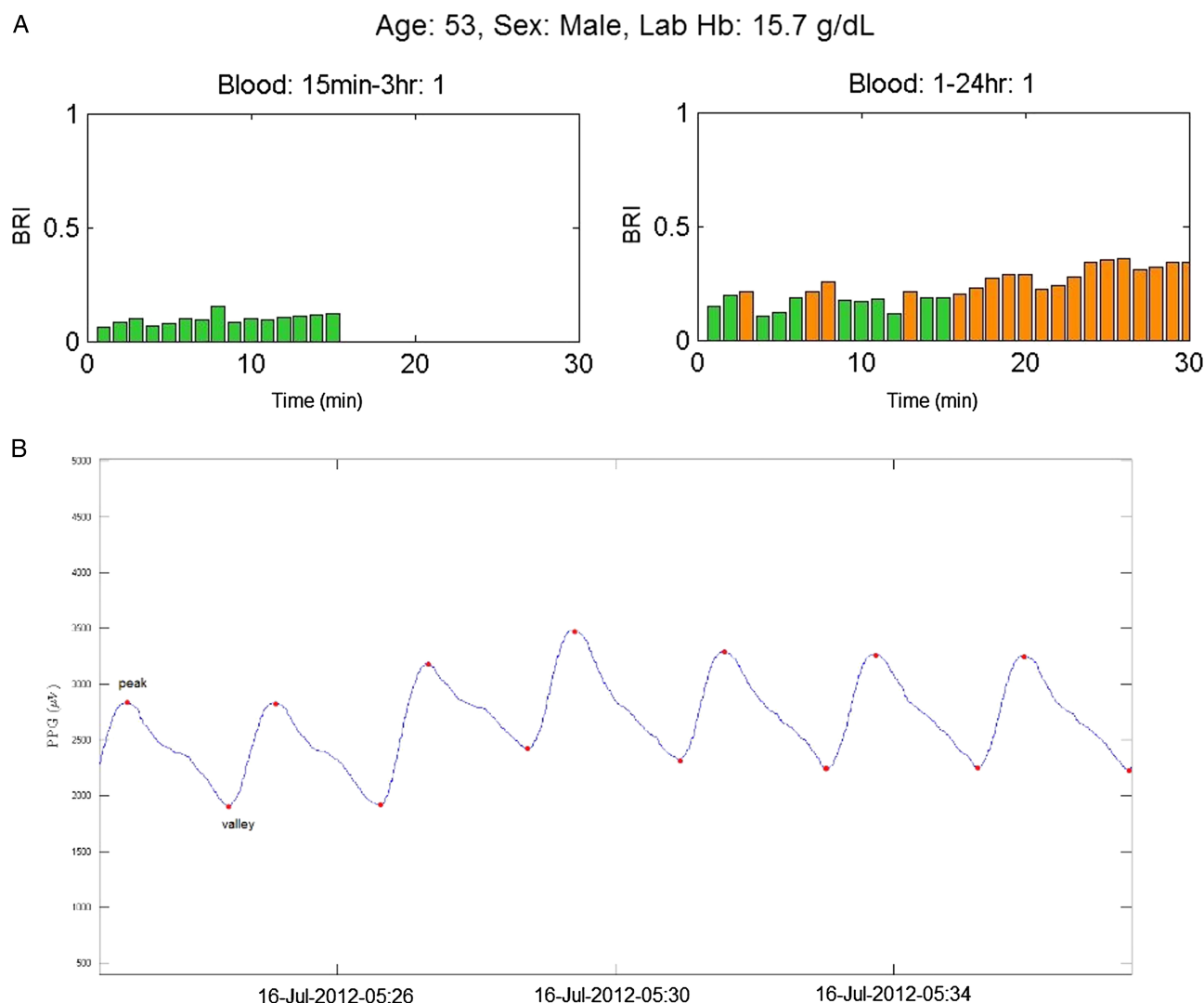


FIG. 2. **A, BRI on the y axis.** Values in green indicate a low probability of transfusion predicted by the pulse oximetry features used in this study, values in yellow indicate a modest probability, and red (not shown in this patient) indicate a high probability of transfusion within the time intervals shown on the x axis. **B,** Ten seconds of the PPG waveform for this patient with amplitude on the y axis and time on the x axis. Peak and valleys of the PPG waveform define the amplitude used in determination of the waveform features (see text).

difference between testing and training of 14% compared with 18% to 21% seen with predictions of the clinicians.

## DISCUSSION

The findings of this preliminary study show that an algorithm derived from features of a single vital sign device can predict the need for surgical interventions and intubation as accurately as the judgment of expert clinicians. In particular, the pulse oximeter-derived algorithm predicted the use of blood transfusion within 6 h of trauma patient admission more accurately than the clinical judgment of expert MDs and PHPs or the majority vote including RNs. This is encouraging because prediction of blood use is a surrogate indicator for hemorrhagic shock. The algorithm was also able to predict patients requiring more than 1,500 mL of fluids within their first hour of resuscitation better than clinical judgment of RNs and PHPs or the majority vote including MDs, confirming previous studies that demonstrated that analysis of pulse oximetry waveforms can detect hypovolemia (14–18). The algorithm predictions for defined surgical interventions and use

of endotracheal intubation were as accurate as expert clinical judgment in this patient cohort.

### **Comparison of PPG with massive transfusion scoring systems**

Traditional massive transfusion scoring systems (6,7,10,19,20) require the input of various diagnostic results such as vital signs, focused assessment with sonography for trauma (FAST) results, radiographic findings, laboratory analyses (hemoglobin, base excess, INR, lactate), and the calculation of a score to predict the need for blood transfusion. Such calculations are rarely useful for clinicians engaged in lifesaving care, and the specific data required for the prediction are not routinely available in most prehospital or austere military environments. In addition, these predictions are based on data collected at a single point in time and do not reflect the dynamic changes that occur in an actively bleeding patient. In comparison, automated analysis of PPG waveforms, although collected after admission to the trauma center in our study, could potentially be applied in any location where pulse oximetry is available, can calculate predictions without user input, and can continuously update to reflect a



TABLE 2. Results of training and testing by leave-one-out methodology showing each intervention, the AUROC by each category of clinicians MDs, RNs, PHPs in comparison with the PPG-derived AUROC and the training and testing difference (diff %)

	PPG data	PHP	RN	MD
<b>Blood</b>				
Training	0.92	0.68	0.71	0.70
Testing	0.78	0.50	0.50	0.50
Difference %	14	18	21	20
<b>Fluid bolus</b>				
Training	0.90	0.62	0.57	0.71
Testing	0.69	0.50	0.50	0.50
Difference %	21	12	7	22
<b>Surgery</b>				
Training	0.74	0.83	0.77	0.65
Testing	0.50	0.68	0.58	0.50
Difference %	24	14	19	15
<b>Intubation</b>				
Training	0.92	0.77	0.90	0.85
Testing	0.65	0.57	0.80	0.72
Difference %	27	19	10	13

patient's changing condition. In reality, a PPG-derived algorithm displayed as a probability of bleeding index (such as Bleeding Risk Index in Fig. 2) would not function independently but would be incorporated into an array of information available to clinicians and would itself become incorporated into the clinicians' judgment. Automated PPG waveform analysis has potential for great clinical utility because of the immediacy of feedback to clinicians with optional electronic communication to the blood bank.

#### Future development of PPG decision support

The next step toward developing a validated decision-support pulse oximeter PPG algorithm will be to compare the clinical judgment of trauma care providers both with and without the assistance of decision-support. Innovative utilization of these findings will support ongoing development of the analytic platforms for future generations of clinical decision-support instrumentation. This approach, although promising, will require further testing in a patient sample with a greater incidence of the outcomes (transfusion, emergency surgical interventions, fluid bolus, and intubation) and a requirement to demonstrate robustness of the selected model by showing AUROC differences of less than 10% during training and testing (13).

#### Previous LSI outcome prediction studies based on vital signs

In previous work by others, features of vital signs have been shown to identify reduced central blood volume in volunteers (17). However, in several studies, the details of the feature extraction from waveforms were not published, although these may be similar to the algorithm developed in our study. The redistribution of blood induced by the lower-body negative pressure model is not accompanied by tissue injury with

mediator release and blood loss as occurs in trauma patients. This negative body pressure-derived algorithm is as of yet untested in trauma patients and may not represent the same physiology seen in blood loss from trauma.

A recent publication tested the Murphy Factor, a composite alert that is based on available vital signs from an 8-oz wireless vital sign sensor and a monitor that collects skin temperature, SpO<sub>2</sub>, HR, and pulse transit time. In 96 prehospital trauma patients of whom 50% had any of six LSIs, a Murphy Factor more than 3 had AUROC of 0.62 for prediction of any LSI. The Murphy Factor did not discriminate which of the six LSIs, one of which was blood transfusion, was indicated (21). In another recent study, HR complexity and variability gave AUROC of 0.81 in comparison with HR alone (AUROC, 0.73) for predicting LSIs in 32 trauma patients, similar to our PPG-based predictions (22).

#### Previous studies assessing clinical judgment

A number of investigators have previously attempted to quantify clinical judgment in various patient care scenarios. Thompson et al. (23) investigated nurses' critical event risk assessment of the need for an intervention in 50 simulated clinical scenarios with and without time pressure and a protocol template. The nurses overestimated the need for intervention in these simulated patients. The authors concluded that unaided decision making may not be as accurate as supported decision making. There are several studies of PHP clinical judgment of triage and diagnosis with variable findings. Mulholland et al. (24) reviewed the literature on triage and found that there is no clear evidence supporting paramedic judgment as an accurate triage method. In a further study of 207 patients, the same authors found that paramedics were unable to reliably identify severe injury to individual body regions; however, the sensitivity for paramedic determination of major trauma was high with correct categorization of all patients admitted to an intensive care unit, those who required urgent surgery, or those who died in the hospital (25). Baxt et al. (26) evaluated 653 prehospital trauma patients using paramedic judgment and the Trauma Triage Rule for accuracy in identifying patients requiring trauma center care. The combination of Trauma Triage Rule and paramedic judgment achieved sensitivity of 100% and specificity of 75% in identifying seriously injured patients (27). We have recently described the findings of 209 clinical judgments of LSI obtained at our trauma center in a similar patient population to those included in the current study and found no "excellent" (Kappa Statistic  $\geq 0.81$ ) agreement between any pair of trauma clinician provider groups (physicians, nurses or pre-hospital providers) for any LSI. The percentage agreement across the different clinical provider groups ranged from 50%–83% (28).

#### Study limitations

Limitations of this study include a small sample size, with skewed data, a limited number of LSI outcomes of interest, and a low incidence of penetrating trauma. The PPG waveform recording during resuscitation in the trauma center has many technical challenges, such as sensor dislodgment, artifact, inability to obtain a signal during poor peripheral perfusion, in

patients with cool extremities, and so on so may limit application of the algorithm to all trauma patient resuscitations. The patient cohort studied was representative of the entire trauma center population with respect to the incidence of blood transfusion of 7.4%; this was within the 6% to 8% incidence of the overall trauma center transfusion rate (29), indicating that the patients studied were similar to that of the overall trauma center population. Not all patients were included because we excluded patients who expired within the first 15 min after admission. In addition, we were unable to survey clinician's predictions when there were concerns with interruption of emergency clinical care, especially in the most severely injured and moribund patients, thus resulting in a potential selection bias to the less injured patient with few LSIs. We did not rank the relative importance of LSIs or the occurrence of multiple LSIs; the pulse oximeter PPG analysis algorithm only considered LSIs one at a time.

The PHPs and RNs were not involved in patient care or implementation of LSIs, and 80% of the MDs completed the clinical judgment survey within 10 minutes and 100% within 15 minutes. The physicians' greatest involvement was in the initial management, hence limiting their ability to complete the survey earlier. In a study by Kim et al. (29), among 245 Level I and II trauma centers in the United States, 82% had trauma surgeons available within 15 min of patient admission.

## CONCLUSIONS

In the patient cohort studied, a pulse oximeter-derived algorithm collected during trauma resuscitation from a single noninvasive vital sign device predicts the need for LSIs with no user input as accurately as experienced clinicians. The algorithm performance suggests that such decision-support may be useful for clinicians with less trauma care training than the experts included in this study. Such automated decision-support could assist resuscitation decisions, trauma team, and operating room and blood bank preparations.

## ACKNOWLEDGMENTS

The authors thank the CORE Research Staff, the clinicians of the R. Adams Cowley Shock Trauma Center, and Betsy Kramer, MS, and Melissa Binder, MS, without whose help the data collection and Trauma Registry review for this study would not have been possible.

## REFERENCES

- Mabry RL, Apodaca A, Penrod J, Orman JA, Gerhardt RT, Dorlac WC: Impact of critical care-trained flight paramedics on casualty survival during helicopter evacuation in the current war in Afghanistan. *J Trauma Acute Care Surg* 73:S32–S37, 2012.
- Gerhardt RT, Cap AP, Cestero R, Dubick MA, Heiner J, Koller AR, Lariet J, McClinton AR, Manfold A, Stewart R, et al.: The Remote Trauma Outcomes Research Network rationale and methodology for study of prolonged out-of-hospital transport intervals on trauma patient outcome. *J Trauma Acute Care Surg* 75:S137–S141, 2012.
- O'Reilly D, Mahendran K, West A, Shirley P, Walsh M, Tai N: Opportunities for improvement in the management of patients who die from hemorrhage after trauma. *Br J Surg* 110:749–755, 2013.
- Dutton RP, Stansbury LG, Leone S, Kramer E, Hess JR, Scalea TM: Trauma mortality in mature trauma systems: are we doing better? An analysis of trauma mortality patterns, 1997–2008. *J Trauma* 69:620–626, 2010.
- Eastridge BJ, Mabry RL, Seguin P, Cantrell K, Tops T, Uribe P, Mallett O, Zubko T, Oetjen-Gerdes L, Rasmussen TE, et al.: Death on the battlefield (2001–2011): implications for the future of combat casualty care. *J Trauma Acute Care Surg* 73:S431–S437, 2012.
- Cotton BA, Au BK, Nunez TC, Gunter OL, Robertson AM, Young PP: Predefined massive transfusion protocols are associated with a reduction in organ failure and post injury complications. *J Trauma* 66:41–48, 2009.
- Nunez TC, Voskresensky IV, Dossett LA, Shinnall R, Dutton WD, Cotton BA: Early prediction of massive transfusion in trauma: simple as ABC (assessment of blood consumption)? *J Trauma* 66:346–352, 2009.
- Mackenzie CF, Wang Y, Hu PF, Chen SY, Chen H, Hagegeorge G, Stansbury L, Shackelford S: Automated prediction of early blood transfusion and mortality in trauma patients. *J Trauma Acute Care Surg* 76(6):1379–1385, 2014.
- Cronbach LJ: Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334, 1951.
- Shackelford SA, Colton K, Stansbury LG, Galvagno SM, Anazodo A, DuBose JJ, Hess JR, Mackenzie CF: Early identification of uncontrolled hemorrhage after trauma: current status and future direction. *J Trauma Acute Care Surg* 77(3 Suppl 2):S222–S227, 2014.
- Shackelford C, Miller K, Frank Y, Wang S, Yang L, Hartsky A, Anazodo J, DuBose R, Fang P, Hu C, et al.: Predicting blood transfusion using automated analysis of pulse oximetry waveforms and lab values. Presented at Military Health Services Research Symposium, August 2014, Ft. Lauderdale, FL.
- DeLong ER, DeLong DM, Clarke-Pearson DL: Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845, 1988.
- Bishop CM: Pattern Recognition and Machine Learning, Springer, New York, 2006, 1–66.
- Loupec T, Nanadoumgar H, Frasca D, Petitpas F, Laksiri L, Baudouin D, Debaene D, Dahyot-Fizelier C, Mimoz O: Pleth variability index predicts fluid responsiveness in critically ill patients. *Crit Care Med* 39:294–299, 2011.
- Cannesson M, Desebbe O, Rosamel P, Delannoy B, Robin J, Bastien O, Lehot J: Plethvariability index to monitor the respiratory variations in the pulse oximeter plethysmographic waveform amplitude and predict fluid responsiveness in the operating theatre. *Br J Anaesth* 101:200–206, 2008.
- Forget P, Lois F, de Kock M: Goal directed fluid management based on pulse-oximeter derived pleth variability index reduces lactate levels and improves fluid management. *Anesth Analg* 111:910–914, 2010.
- Partridge BL: Use of pulse oximetry as a non-invasive predictor of intravascular volume status. *J Clin Monit* 3:236–238, 1987.
- Moulton SL, Mulligan J, Grudic GZ, Convertino VA: Running on empty? The compensatory reserve index. *J Trauma Acute Care Surg* 77:1053–1059, 2013.
- Van Haren RM, Thorson MC, Valle EJ, Busko AM, Jouria JM, Livingstone AS, Namias N, Schulman CI, Proctor KG: Novel prehospital monitor with injury acuity alarm to identify trauma patients who require lifesaving intervention. *J Trauma Acute Care Surg* 76:743–749, 2014.
- Ruchholtz S, Pehle B, Lewan U, Lefering R, Muller N, Oberbeck R, Waydhas C: The emergency room transfusion score (ETS): prediction of blood transfusion requirement in initial resuscitation after severe trauma. *Transfus Med* 16:49–56, 2006.
- Yücel N, Lefering R, Maegele M, Vorweg M, Tjardes T, Ruchholtz S, Neugebauer EA, Wappler F, Bouillon B, Rixen D, et al.: Trauma Associated Severe Hemorrhage (TASH)- Score: probability of mass transfusion as surrogate for life threatening hemorrhage after multiple trauma. *J Trauma* 60:1228–1237, 2006.
- Liu NT, Holcomb JB, Wade CE, Darrah MI, Salinas J: Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for life-saving interventions in trauma patients. *Shock* 42:108–114, 2014.
- Thompson C, Bucknall T, Estabrookes CA, Hutchinson A, Fraser K, de Vos R, Binneade J, Barrat G, Saunders J: Nurses' critical event risk assessments: a judgment analysis. *J Clin Nurs* 18:601–612, 2009.
- Mulholland SA, Gabbe BJ, Cameron P: Victoria State Outcomes Registry and Monitoring group (VSTORM). Is paramedic judgment useful in prehospital trauma triage? *Injury* 36:1298–1305, 2005.
- Mulholland SA, Cameron P, Gabbe BJ, Williamson OD, Young K, Smith KL, Bernard SA: Prehospital prediction of the severity of blunt anatomic injury. *J Trauma* 64:740–760, 2008.
- Fries GR, McCalla G, Levitt MA, Cordova R: A prospective comparison of paramedic judgment and the trauma triage rule in the prehospital setting. *Ann Emerg Med* 24:885–889, 1994.
- Baxt WG, Jones G, Fortlage D: The trauma triage rule: a new resource-based approach to pre-hospital identification of major trauma victims. *Ann Emerg Med* 19:1401–1406, 1990.
- de Biasi AR, Stansbury LG, Dutton RP, Stein DM, Scalea TM, Hess JR: Blood product use in trauma resuscitation: plasma deficit versus plasma ratio as predictors of mortality in trauma. *Transfusion* 51:1925–1932, 2011.
- Anazodo AN, Murthi S, Frank MK, Hu P F-M, Hartsky L, Imle PC, Stephens C, Menaker J, Miller C, DiNardo T, et al.: Assessing Trauma-Care Provider Judgment in the Prediction of Need for Life-Saving Interventions. *Injury*. Epub 2014/11/18. doi:10.1016/j.injury.2014.10.063. PubMed PMID: 25541418.
- Kim Y-J, Xiao Y, Mackenzie CF, Gardner S: Availability of trauma specialists in level I and II trauma centers: a national survey. *J Trauma* 63(3):676–683, 2007.